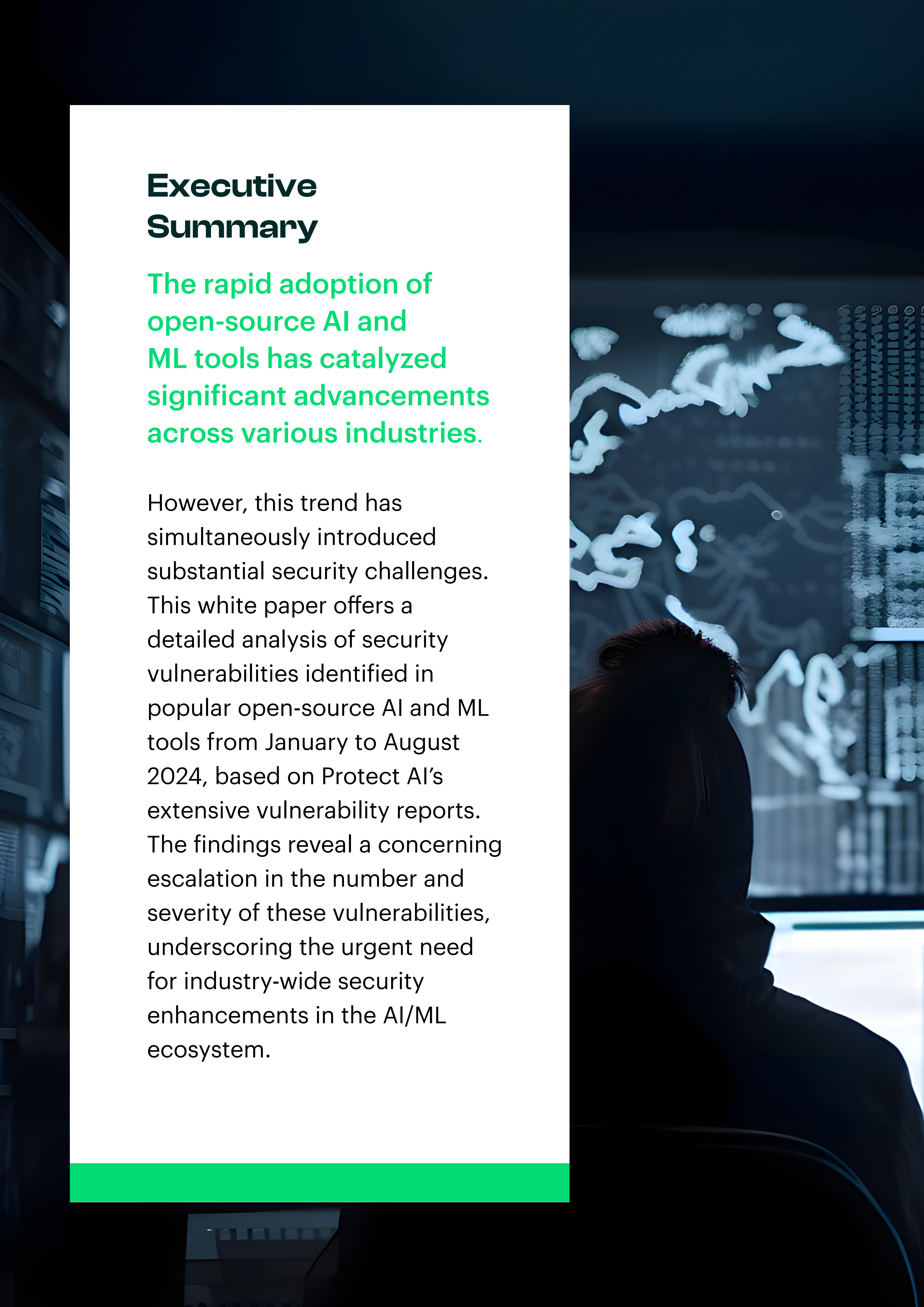
AI Cyber Insights

Industry Implications of **Security Vulnerabilities in Open Source AI and ML Tools**



Executive Summary

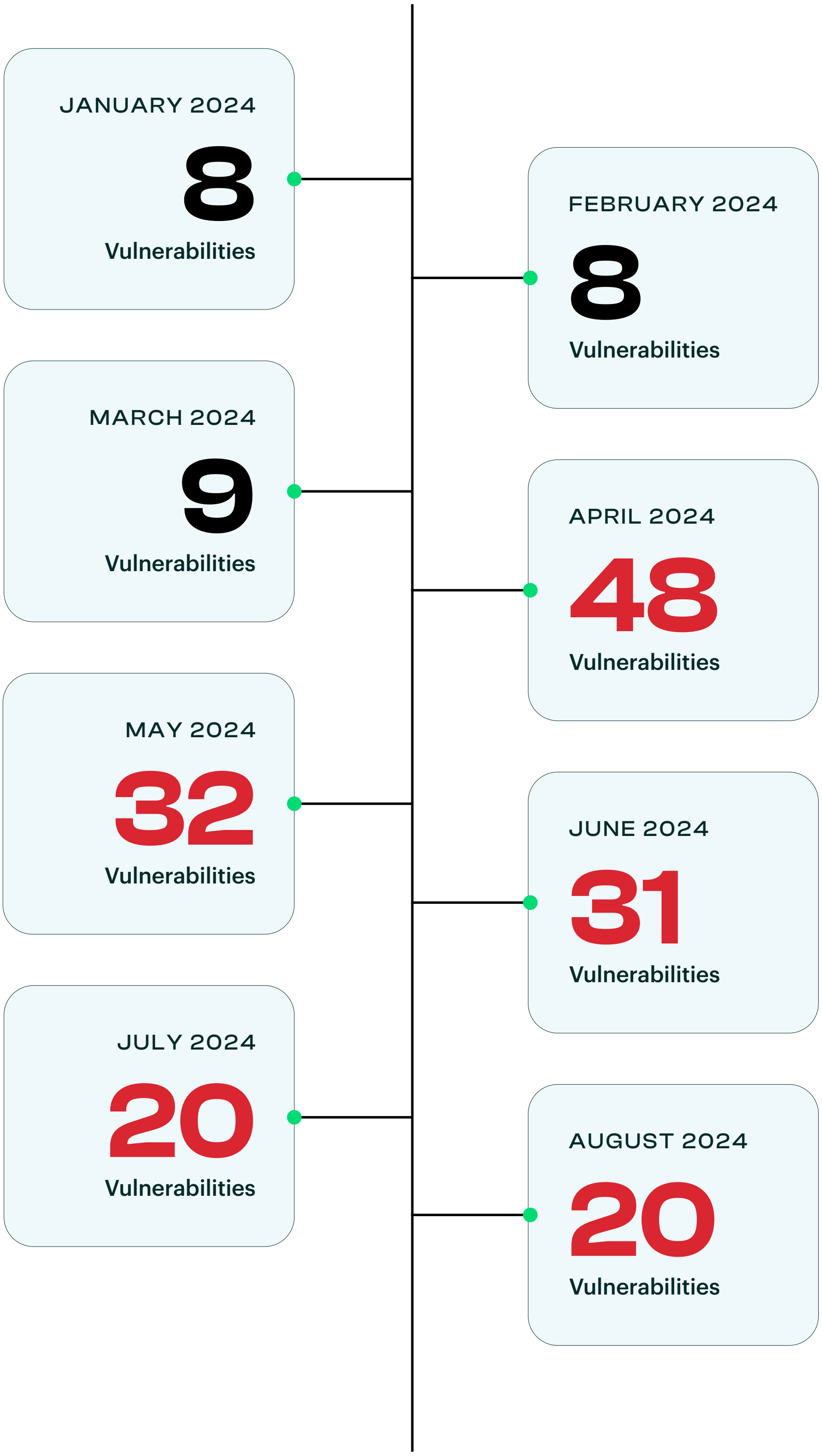
The rapid adoption of open-source AI and ML tools has catalyzed significant advancements across various industries.

However, this trend has simultaneously introduced substantial security challenges. This white paper offers a detailed analysis of security vulnerabilities identified in popular open-source AI and ML tools from January to August 2024, based on Protect AI's extensive vulnerability reports. The findings reveal a concerning escalation in the number and severity of these vulnerabilities, underscoring the urgent need for industry-wide security enhancements in the AI/ML ecosystem.

Key Insights

Vulnerability Growth and Distribution:

Between January and August 2024, a total of 176 vulnerabilities were publicly disclosed across various open-source AI and ML tools. A monthly breakdown of these vulnerabilities shows a disturbing upward trend, peaking in April 2024 with 48 vulnerabilities—marking the largest publication to date.



JANUARY 2024
8
Vulnerabilities

FEBRUARY 2024
8
Vulnerabilities

MARCH 2024
9
Vulnerabilities

APRIL 2024
48
Vulnerabilities

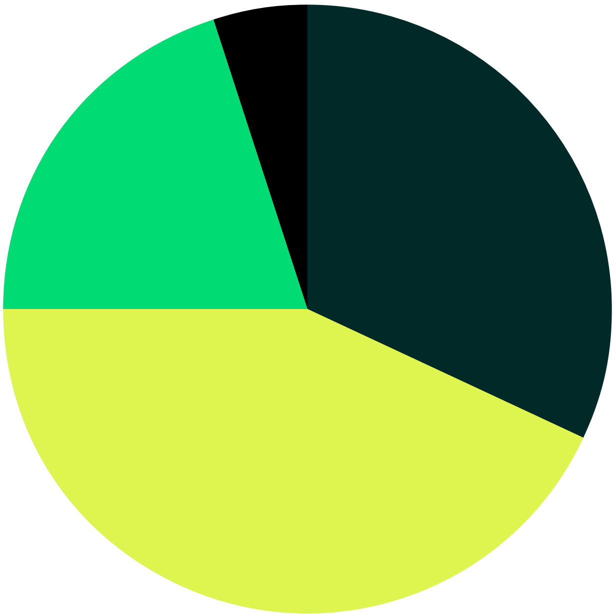
MAY 2024
32
Vulnerabilities

JUNE 2024
31
Vulnerabilities

JULY 2024
20
Vulnerabilities

AUGUST 2024
20
Vulnerabilities

The severity of these vulnerabilities is particularly alarming:



- ■ Critical: 32% (56 vulnerabilities)
- 🟨 High: 43% (75 vulnerabilities)
- 🟩 Medium: 20% (35 vulnerabilities)
- 🟩 Low: 5% (9 vulnerabilities)

This distribution highlights the significant risks these vulnerabilities pose, with the majority classified as Critical or High, indicating a high potential for exploitation if left unaddressed.

Most Affected Tools:

Several widely used AI/ML tools have been disproportionately affected:

MLflow: 15 reported vulnerabilities

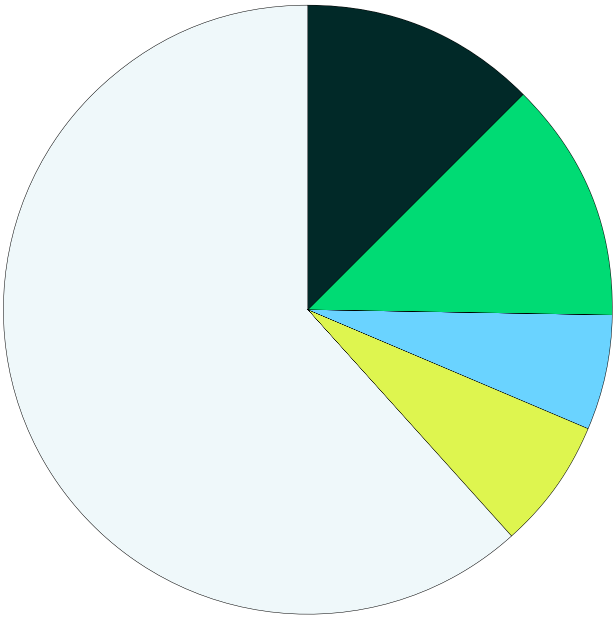
anything-llm: 33 reported vulnerabilities

lollms: 22 reported vulnerabilities

These tools are integral to various AI/ML workflows, collectively accounting for 40% of all reported issues, making them critical targets for security enhancements.

Common Vulnerability Types:

The reported vulnerabilities exhibit a diverse range of attack vectors, with the most prevalent being:



- ■ Remote Code Execution (RCE): 12.5% (22 vulnerabilities)
- 🟩 Path Traversal: 13% (23 vulnerabilities)
- 🟦 Privilege Escalation: 6% (11 vulnerabilities)
- 🟨 Server-Side Request Forgery (SSRF): 7% (12 vulnerabilities)
- □ Other Vulnerabilities

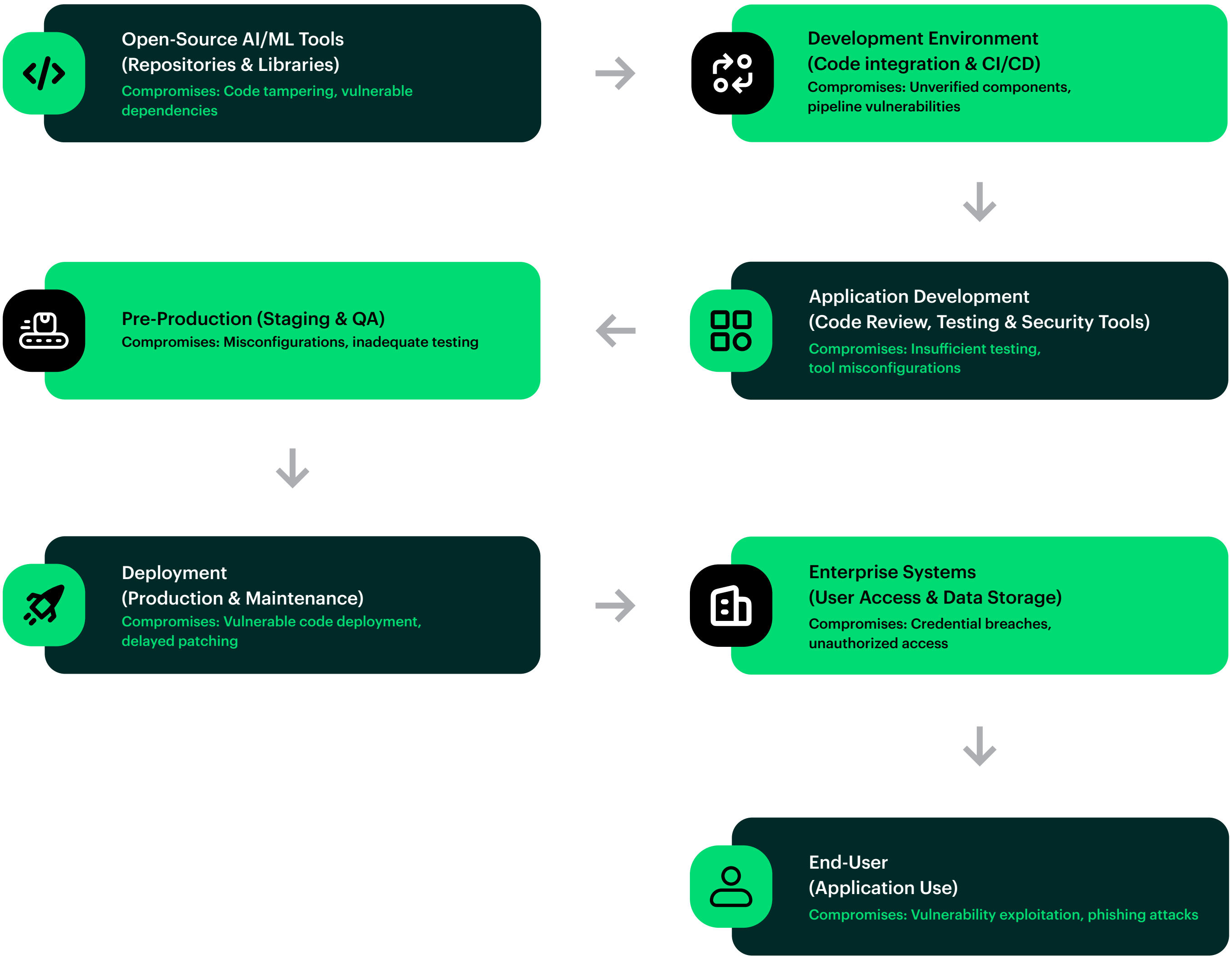
These attack vectors represent critical entry points for attackers, allowing them to execute arbitrary code, gain unauthorized access, or escalate privileges, potentially compromising entire systems.

Industry Impact Analysis

The vulnerabilities identified pose significant risks across the AI/ML industry:

Software Supply Chain Risks:

The widespread use of affected tools like MLflow, Triton Inference Server, and PyTorch in AI/ML development pipelines means that vulnerabilities could lead to extensive infections across numerous projects and organizations.



Open-Source AI/ML Tools
(Repositories & Libraries)
Compromises: Code tampering, vulnerable dependencies

Development Environment
(Code integration & CI/CD)
Compromises: Unverified components, pipeline vulnerabilities

Application Development
(Code Review, Testing & Security Tools)
Compromises: Insufficient testing, tool misconfigurations

Pre-Production (Staging & QA)
Compromises: Misconfigurations, inadequate testing

Deployment
(Production & Maintenance)
Compromises: Vulnerable code deployment, delayed patching

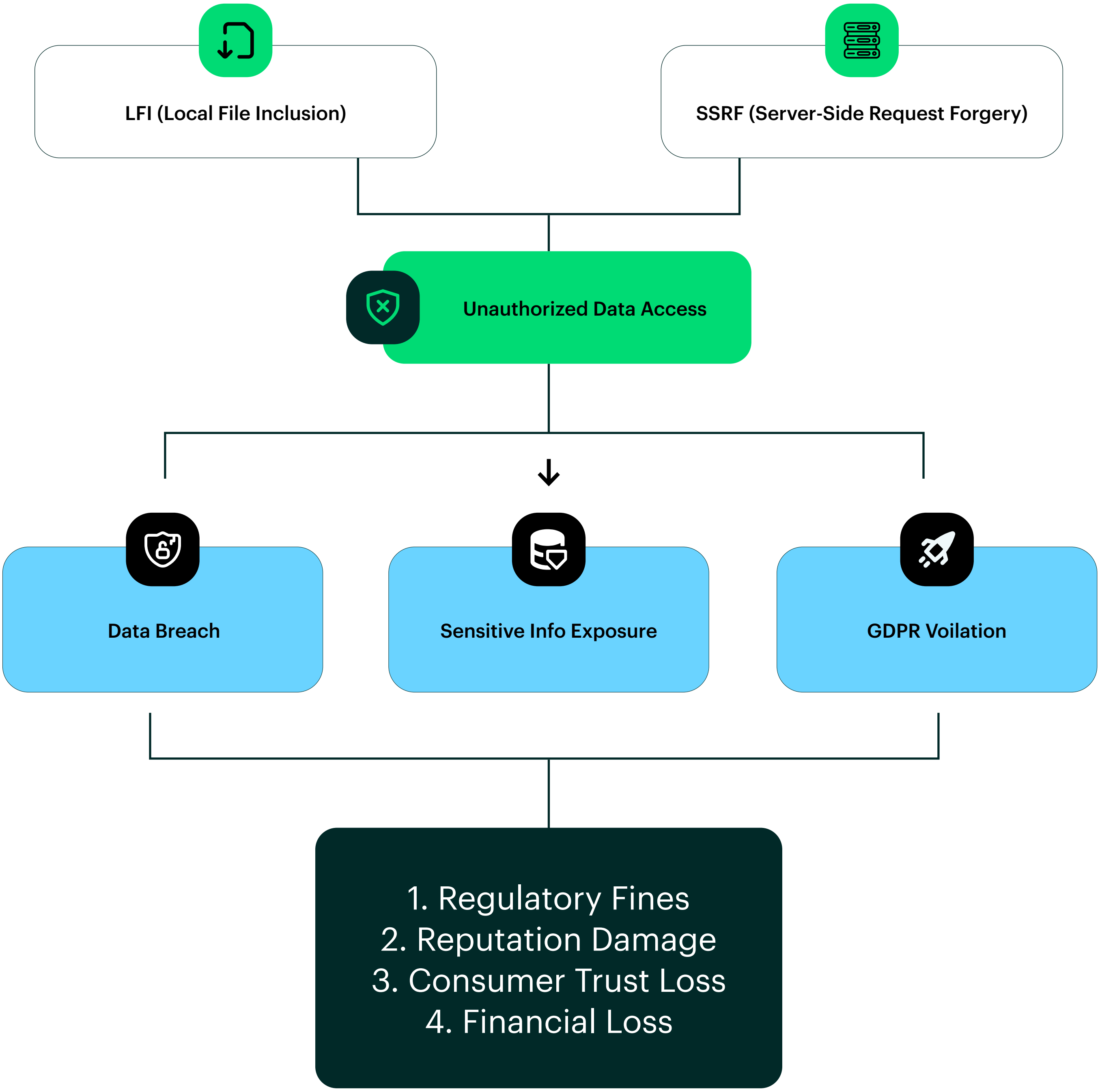
Enterprise Systems
(User Access & Data Storage)
Compromises: Credential breaches, unauthorized access

End-User
(Application Use)
Compromises: Vulnerability exploitation, phishing attacks

SOFTWARE SUPPLY CHAIN RISKS

Data Privacy Concerns:

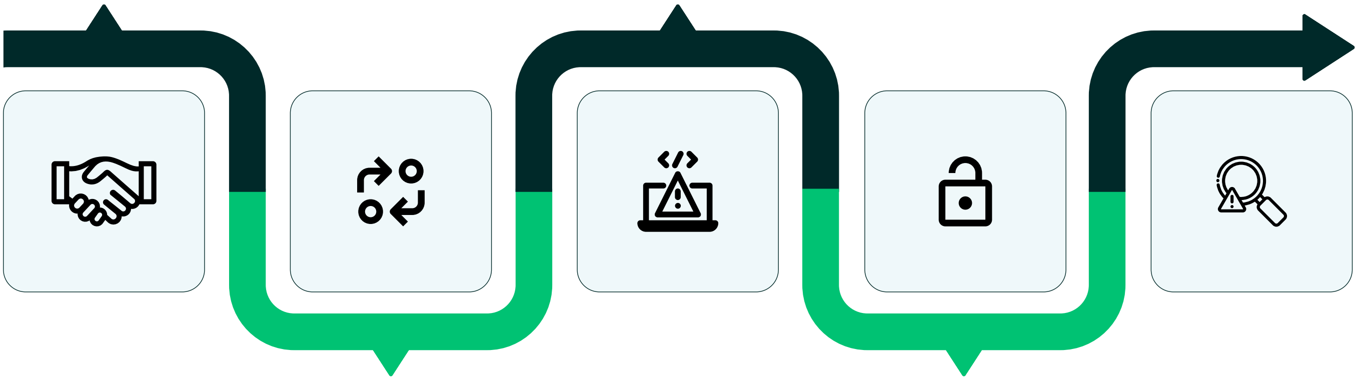
Vulnerabilities such as Local File Inclusion (LFI) and SSRF can lead to unauthorized access to sensitive data, potentially resulting in violations of data protection regulations, such as the General Data Protection Regulation (GDPR).



VULNERABILITIES IN AI COULD LEAD TO PRIVACY VIOLATIONS.

Model Integrity:

Vulnerabilities allowing arbitrary file writes or RCE could enable attackers to tamper with ML models, potentially introducing backdoors or biases, thereby compromising the integrity of AI models.



1. Introduction of Vulnerability

SOURCE:
Open source repository or third-party model.

COMPROMISES:
Code tampering, vulanerable dependencies.

2. Model Integration

ENVIRONMENT:
Development or production.

ACTION:
Tampered model is integrated into an AI/ML system without proper validation.

3. Exploitation of Vulnerability

TRIGGER:
Malicious actors exploit the tampered mode.

ACTION:
Unauthorized access, data manipulation, or triggering unintended behaviors in the system

4. Compromised AI Outputs

OUTCOME:
Biased predictions, incorrect classifications, or manipulated decisions.

IMPACT:
Negative consequences on business operations, security breaches, or loss of trust.

5. Detection and Mitigation

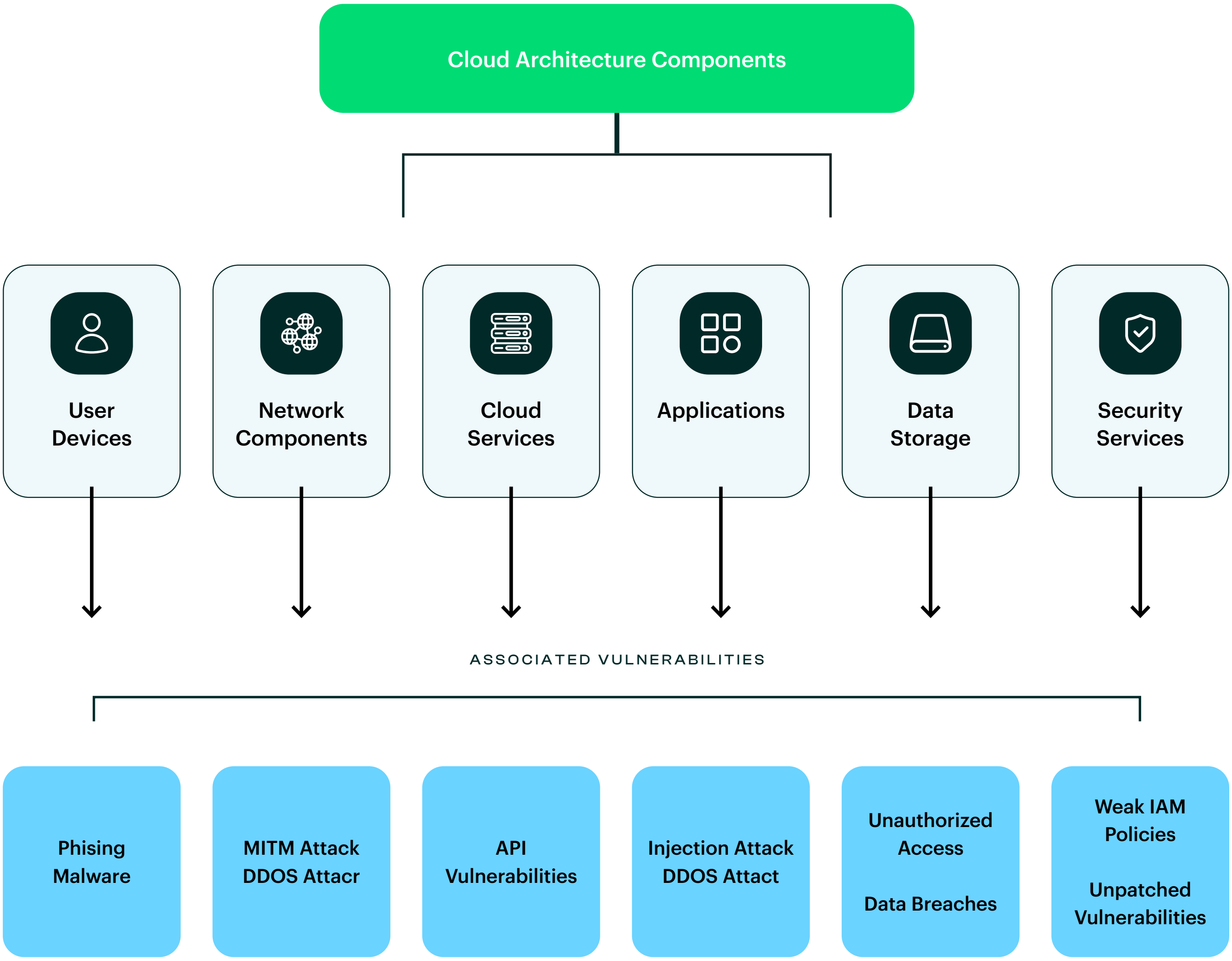
ACTION:
Post-deployment analysis detects anomalies

RESPONSE:
Patching vulnerabilities, retraining models, or rolling back to a secure version

ILLUSTRATION THE IMPACT OF COMPROMISED ML MODELS

Cloud Infrastructure Vulnerabilities:

Several vulnerabilities, particularly in tools like MLflow and Triton Inference Server, affect tools often deployed in cloud environments, increasing the risk of broader cloud infrastructure compromises.



THIS SKETCH ABOVE SHOWS A CLOUD ARCHITECTURE DIAGRAM REPRESENTING DIFFERENT COMPONENTS OF THE CLOUD ENVIRONMENT

Financial Implications

The potential for industry-wide financial impact is significant, with the risk of exploitation reaching into the billions if these vulnerabilities are not mitigated.

Interconnected Vulnerability Landscape

1. Credential Harvesting Chain

- SSRF vulnerabilities (7% of total) often lead to exposure of cloud metadata
- Exposed metadata enables privilege escalation (6% of vulnerabilities)
- Elevated privileges facilitate large-scale data exfiltration

2. Model Poisoning Pipeline

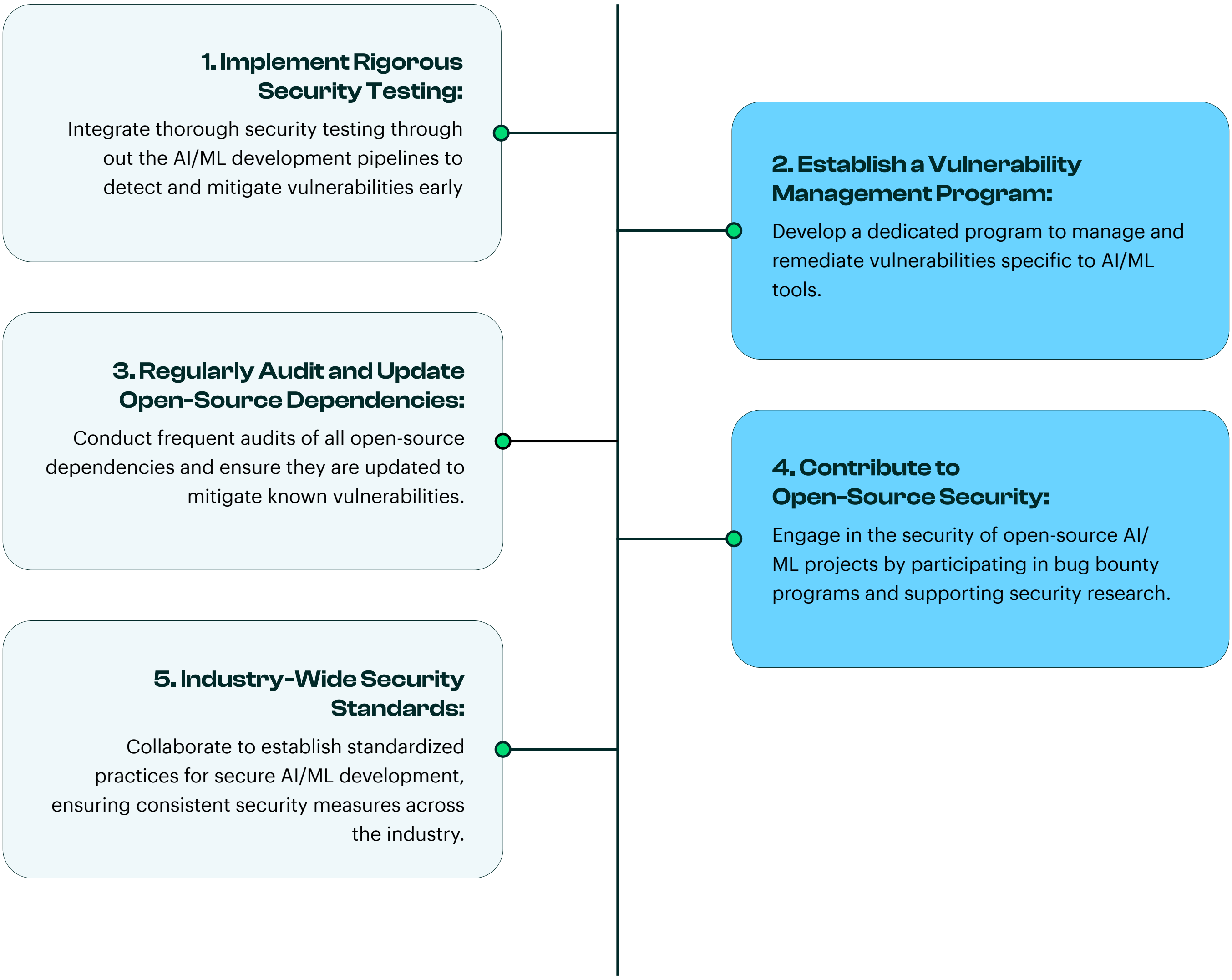
- Path traversal issues (13% of total) allow access to model storage
- Unauthorized access enables model tampering
- Compromised models deployed widely, affecting downstream applications

3. Distributed Denial of Service (DDoS) Amplification

- Vulnerabilities allowing arbitrary file operations or RCE in popular libraries could be exploited for reflection and amplification attacks.

Recommendations

To address these security challenges, the following measures are recommended:



1. Implement Rigorous Security Testing:

Integrate thorough security testing through out the AI/ML development pipelines to detect and mitigate vulnerabilities early

2. Establish a Vulnerability Management Program:

Develop a dedicated program to manage and remediate vulnerabilities specific to AI/ML tools.

3. Regularly Audit and Update Open-Source Dependencies:

Conduct frequent audits of all open-source dependencies and ensure they are updated to mitigate known vulnerabilities.

4. Contribute to Open-Source Security:

Engage in the security of open-source AI/ML projects by participating in bug bounty programs and supporting security research.

5. Industry-Wide Security Standards:

Collaborate to establish standardized practices for secure AI/ML development, ensuring consistent security measures across the industry.

Conclusion

The discovery of 176 vulnerabilities across major AI/ML tools within just eight months highlights a critical security challenge for the AI industry. With 75% of these vulnerabilities rated as Critical or High severity, the potential for exploitation is deeply concerning. The interconnected nature of these vulnerabilities creates a complex risk landscape that extends beyond individual tools or organizations. As AI continues to permeate critical systems and decision-making processes, addressing these security challenges is crucial for the responsible and safe advancement of AI technologies.

Contributors

Lead author

Confidence Staveley
Editor-In-Chief – AI Cyber Insights

Co-authors

Isu Abdulrauf
AI Researcher – AI Cyber Insights

Victoria Robinson
AI Researcher – AI Cyber Insights

Additional Reading

1. https://www.linkedin.com/pulse/rise-open-source-ai-david-cain-gymoc/
2. https://protectai.com/threat-research
3. https://protectai.com/threat-research/january-vulnerability-report
4. https://protectai.com/threat-research/february-vulnerability-report
5. https://protectai.com/threat-research/march-vulnerability-report
6. https://protectai.com/threat-research/april-vulnerability-report
7. https://protectai.com/threat-research/may-vulnerability-report
8. https://protectai.com/threat-research/june-vulnerability-report
9. https://protectai.com/threat-research/july-vulnerability-report
10. https://protectai.com/threat-research/august-vulnerability-report

www.aicyberinsights.com